Proceedings of ICFTBA 2026 Symposium: Digital Government, Finance, Accountability, and Sustainable Economic Development
DOI: 10.54254/2754-1169/2026.AB38131

Analysis and Forecasting of FTSE 100 Return Volatility Using an ARMA-GARCH Model

Jiaquan Wang

University of Bristol, Bristol, UK
13660771169@163.com

Abstract. This paper investigates whether a relatively simple ARMA-GARCH specification is useful for modelling and forecasting daily FTSE 100 returns. Closing prices from 31 May 2023 to 28 August 2026 generate 823 percentage log returns. Of these, 758 observations are used for estimation and the final 65 trading days are kept aside as an out-of-sample test. Several low-order ARMA and GARCH specifications, under normal and Student-t errors, are compared using only the estimation sample. BIC selects ARMA(0,0) for the mean equation. For volatility, Student-t GARCH(1,1) is retained because it is parsimonious and its parameters are statistically significant. Out of sample, the mean forecast gives an RMSE of 0.5760 and an MAE of 0.4190, only slightly better than a zero-return benchmark. The reported directional accuracy of 56.9% is largely explained by the positive constant mean, which produces a positive sign forecast on every test day. For conditional variance, GARCH performs better under MAE, whereas the 20-day rolling variance has lower MSE and QLIKE. Overall, the model is more useful for tracking changes in risk than for forecasting the short-run direction of FTSE 100 returns.

Keywords: FTSE 100, ARMA, Student-t GARCH, conditional volatility, out-of-sample forecasting

1. Introduction

Volatility in a stock index changes with economic news, company performance and shifts in investor expectations. It also matters in practical decisions such as portfolio allocation, risk control, derivative pricing and capital management. The FTSE 100 is particularly exposed to both domestic and international influences because many of its large listed companies earn a substantial share of their revenues outside the United Kingdom. Daily returns can therefore react to UK interest rates, inflation and fiscal policy as well as to energy prices, global trade conditions and geopolitical events. These return movements are not evenly distributed over time. Large changes often arrive in clusters, with quieter periods in between, which is the familiar pattern of volatility clustering.

That feature makes it sensible to treat the return mean and the return variance as separate modelling problems. An autoregressive moving-average (ARMA) model looks for linear dependence in the mean of returns, whereas a generalized autoregressive conditional heteroskedasticity (GARCH) model describes how volatility changes after earlier shocks and periods of high or low variance. Used together, the two models allow expected return and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.AB38131

conditional risk to be studied in the same framework. Even so, a model that fits the estimation sample well is not necessarily useful for forecasting. If the same observations are used both to choose a model and to judge it, the reported forecasting performance may be too optimistic.

To avoid that problem, model selection is carried out on a fixed estimation sample and the final three months are left untouched until the forecasting stage. One-step-ahead return forecasts are assessed against a zero-return benchmark, while variance forecasts are compared with a 20-day rolling historical variance. The distinction matters because the two parts of the model do different jobs: ARMA supplies the return point forecast and GARCH supplies the conditional variance forecast. The empirical question is therefore not simply whether ARMA-GARCH fits the data, but whether it adds anything useful beyond these simple benchmarks.

2. Literature review

The starting point for modern conditional-variance modelling is the ARCH model, where current variance depends on earlier squared shocks [1]. Bollerslev extended this idea by allowing lagged conditional variance to enter the equation, giving the GARCH model a compact way to represent volatility persistence [2]. Later studies relaxed the assumption that positive and negative shocks have the same effect. Nelson introduced EGARCH [3], while Glosten, Jagannathan and Runkle proposed the GJR specification [4]. One implication of this literature is important for the present paper: a return series may show little predictable movement in its mean and still have a variance process with clear dynamics.

Forecast evaluation became a separate concern in later research. Brailsford and Faff found that the ranking of volatility models depends partly on the loss function used to evaluate them [5]. Andersen and Bollerslev argued that standard volatility models can look weaker than they really are when the target used for evaluation is itself noisy; more informative volatility measures can change that assessment [6]. In work on the UK market, McMillan, Speight and ap Gwilym compared several forecasting approaches at daily, weekly and monthly frequencies. Their results suggest that moving-average and GARCH methods can perform reasonably consistently, although their relative advantage changes with the sampling frequency [7]. Poon and Granger reached a similar broad conclusion in their review: no single forecasting method is best in every market or under every evaluation criterion [8].

Other studies reinforce the same point from different angles. Awartani and Corradi examined whether allowing for asymmetric volatility improves forecasting performance [9]. Hansen and Lunde compared a large number of candidate models and showed that many complicated specifications fail to beat the standard GARCH(1,1) model consistently [10]. Patton focused on the evaluation problem itself, showing why forecast comparisons require care when latent variance is judged using an imperfect observable proxy [11].

GARCH models have also been used widely in Chinese empirical finance. Liu compared GARCH, QGARCH and GJR models for Chinese stock-market volatility and reported that model performance changed once the effect of extreme observations was taken into account [12]. Ye combined GARCH with semiparametric methods in a value-at-risk application [13]. Huang and Wang showed, using the Shanghai Composite Index, that estimation choices can affect the conclusions drawn from a GARCH model [14]. Yang and Cao later applied an ARMA-GARCH framework to stock-price analysis and forecasting, reflecting a broader shift from modelling variance alone to modelling the conditional mean and variance jointly [15].

The literature therefore gives strong reasons to use GARCH when volatility clustering is present, but much weaker grounds for assuming that a fitted model will forecast well out of sample. Recent

DOI: 10.54254/2754-1169/2026.AB38131

daily FTSE 100 data are not a major focus of the Chinese applied literature, and some empirical papers give more attention to in-sample fit or comparisons among competing models than to performance against very simple benchmarks. This paper addresses that gap by keeping estimation and testing separate and by treating return forecasts and variance forecasts as two different empirical questions.

3. Methodology

3.1. ARMA mean equation

Let r_t denote the percentage log return on trading day t. An ARMA(p,q) model represents the conditional mean using past returns and past innovations

$$r_t = c + \Sigma_{i=1}^{p} \varphi_i r_{t-i} + \varepsilon_t + \Sigma^{q}_{j=1} \theta_j \varepsilon_{t-j} \quad (1)$$

In equation (1), c is the constant, while φ_i and θ_j are the autoregressive and moving-average coefficients. The candidate values of p and q run from 0 to 3. Schwarz's Bayesian information criterion (BIC) is used as the main rule for selecting the order [16], while the model-complexity trade-off associated with Akaike is also considered [17]. With ARMA(0,0), there are no lagged AR or MA terms, so the conditional mean reduces to a constant.

3.2. Student-t GARCH variance equation

When the variance of the mean-equation innovation changes through time, the residual can be written as:

$$\varepsilon_t = \sigma_t z_t \quad (2)$$

where z_t follows a standardized Student-t distribution. The GARCH(m,n) conditional variance is specified as:

$$\sigma_t^2 = \omega + \Sigma_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2 + \Sigma_{j=1}^{n} \beta_j \sigma_{t-j}^2 \quad (3)$$

In this equation, ω is the variance constant, α measures the response of volatility to new shocks and β captures persistence from earlier conditional variance. For GARCH(1,1), the usual covariance-stationarity condition requires non-negative coefficients and $\alpha+\beta<1$. A Student-t error distribution introduces a degrees-of-freedom parameter ν and allows more probability in the tails than the normal distribution, which is useful for daily equity returns with occasional large observations.

3.3. Out-of-sample evaluation

Return forecasts are judged by root mean squared error (RMSE), mean absolute error (MAE) and directional accuracy, with a zero-return forecast used as the benchmark. These measures apply to the point forecast from the ARMA mean equation. Variance is different because the true conditional variance is not observed directly. Squared returns r_t^2 are therefore used as a proxy, and GARCH forecasts are compared with a 20-day historical rolling variance using mean squared error (MSE), MAE and QLIKE. MSE reacts strongly to large forecast errors, whereas MAE reflects the typical

DOI: 10.54254/2754-1169/2026.AB38131

absolute error more directly. QLIKE is commonly used for positive variance forecasts and is relatively robust to scale. Since squared returns are noisy measures of latent variance, the three loss functions are read together rather than relying on a single ranking.

4. Data and empirical results

4.1. Data source and sample split

The data are daily closing levels of the FTSE 100 price index, reported under the identifiers UKX or ^FTSE. The earlier observations come from the downloadable FTSE 100 Historical Data file on Investing.com [18]. Later 2026 observations are supplemented and checked against MarketWatch historical quotes for UKX. MarketWatch states that its historical and end-of-day data are supplied by FactSet [19]. The two sources overlap for 22 trading days in March 2026. Across those dates, the largest difference in the reported closing level is only 0.05 index points, which is consistent with rounding. Both sources report a close of 10,824.26 on 28 August 2026, the final day in the sample. The data were checked on 1 September 2026. Weekends and market holidays are left as non-trading days rather than filled or interpolated.

The full price series covers 31 May 2023 to 28 August 2026 and contains 824 trading-day observations, from which 823 daily returns are calculated. The first closing price is needed only to construct the first return. There are 758 returns in the estimation period, giving the volatility model about three years of information. This longer window reduces the chance that parameter estimates are driven mainly by a short episode of unusual market behaviour. The remaining 65 returns, roughly three months of trading, are reserved for out-of-sample testing and are not used for model selection. Returns are multiplied by 100 before estimation, so the return-related coefficients and forecast errors are reported in percentage points (see Table 1):

$$r_t = 100 \times [ln(P_t) - ln(P_{t-1})] \tag{4}$$

Table 1. Sample split

Purpose	Date range	Return observations
Estimation period	2023-06-01 to 2026-05-29	758
Out-of-sample test	2026-06-01 to 2026-08-28	65
Total	2023-06-01 to 2026-08-28	823

4.2. Descriptive statistics and preliminary tests

The mean daily return during estimation is 0.0442%, compared with a standard deviation of 0.7137%. In other words, ordinary day-to-day variation is much larger than the average return. Kurtosis is 9.6892 and the Jarque-Bera test rejects normality, which supports the use of Student-t errors. The augmented Dickey-Fuller test does not reject a unit root in the index level but does reject one in the return series. ARMA modelling is therefore applied to returns rather than to the price level. ARCH-LM tests with five and twelve lags are both significant, giving direct evidence that conditional variance changes over time (see Table 2).

DOI: 10.54254/2754-1169/2026.AB38131

Table 2. Estimation-period return statistics and tests

Measure	Result	Interpretation
Mean	0.0442%	Close to zero
Standard deviation	0.7137%	Daily variation is substantial
Kurtosis	9.6892	Strongly leptokurtic / heavy-tailed
Jarque-Bera	$p<0.001$	Normality rejected
Closing-price ADF	$p=0.9554$	Price unit root not rejected
Return ADF	$p<0.001$	Returns are stationary
ARCH-LM (lags 5 and 12)	both $p<0.001$	Significant ARCH effects

4.3. Model selection

Model choice uses only the 758 estimation-period returns. For the mean equation, all 16 combinations with p and q between 0 and 3 are considered. ARMA(0,0) has the lowest BIC. Adding AR or MA terms improves the fit too little to justify the extra parameters under this criterion. This is consistent with the well-known difficulty of finding stable linear predictability in daily equity-index returns. It says nothing, however, about whether the variance is constant (see Table 3).

Table 3. Model selection criteria

Stage	Candidate specifications	Main evidence	Selected specification
Mean equation	ARMA(p,q), p,q=0,...,3	ARMA(0,0) has the lowest BIC; no clear residual linear dependence	ARMA(0,0)
Innovation distribution	Normal, Student-t	Student-t specifications generally have lower BIC and match the heavy-tailed sample	Student-t
Variance order	GARCH(1,1), (1,2), (2,1), (2,2)	t-GARCH(1,2) lowers BIC by only about 0.78; the extra β term is insignificant	GARCH(1,1)

Note: When BIC differences are very small, parameter significance, parsimony and interpretability are considered alongside the information criterion.

The variance equation is next compared across four low-order GARCH structures, each estimated with normal and Student-t innovations. Student-t versions are generally preferred by BIC. The lowest BIC belongs to t-GARCH(1,2), but the difference from t-GARCH(1,1) is only about 0.78, and the extra lagged-variance coefficient is not statistically significant. Keeping that term would make the model more complicated for very little gain in fit. For that reason, the final specification is ARMA(0,0) with Student-t GARCH(1,1).

4.4. Parameter estimates and diagnostics

ARMA(0,0) has no AR or MA lags, so its conditional mean is just the estimated constant. Using the 758 estimation observations gives $c=0.0442\%$. Consequently, the return point forecast in the test period is 0.0442% every day rather than a value that moves with past returns. Conditional variance is estimated separately using Student-t GARCH(1,1); the parameter estimates are shown in Table 4.

DOI: 10.54254/2754-1169/2026.AB38131

Table 4. Student-t GARCH(1,1) variance parameter estimates

Parameter	Estimate	Std. error	z statistic	p value
ω	0.166268	0.062761	2.649	0.0081
α	0.190572	0.069170	2.755	0.0059
β	0.447363	0.176292	2.538	0.0112
ν	6.0207	1.3777	—	—

Note: Returns are modelled in percentage points, so ω is measured in percentage points squared.

The estimates of ω, α and β are all significant at the 5% level, so both recent shocks and earlier conditional variance help explain next-period volatility. Their sum, α+β=0.6379, is below one and satisfies the usual covariance-stationarity condition. This implies persistence, but not permanence: the effect of a volatility shock fades over time. The estimated degrees of freedom is about six, which is also consistent with a heavy-tailed error distribution. The persistence measure α+β refers to volatility only and should not be confused with the accuracy of predicting return direction.

The fitted model is then checked for dynamics left in the residuals. Standardized residuals are used to look for remaining linear dependence, while squared standardized residuals and ARCH-LM tests are used to examine remaining conditional heteroskedasticity. At the 5% level, the relevant tests do not reject the null hypotheses of no remaining correlation and no remaining ARCH effect. This suggests that the final specification captures the main in-sample patterns identified in the estimation data (see Table 5).

Table 5. Model diagnostic summary

Diagnostic object	Purpose	Decision at 5% level
Standardized residuals	Check remaining linear autocorrelation	No significant remaining correlation
Squared standardized residuals	Check remaining volatility dependence	No significant remaining correlation
ARCH-LM	Check remaining conditional heteroskedasticity	No significant remaining ARCH effect

These diagnostic results are useful, but they are not a forecast test. A model can leave little obvious structure in its in-sample residuals and still perform poorly on new data. Daily financial series are especially exposed to news shocks, changing market conditions and noise in the proxy used for volatility. The separate test period is therefore needed to judge whether the fitted model adds predictive value.

5. Out-of-sample forecasting results

5.1. Forecast design

The ARMA order, GARCH order, innovation distribution and parameter estimates are fixed using information available up to 29 May 2026. The specification is not reselected or re-estimated using future test-period observations. Because the selected mean model is ARMA(0,0), the one-step return forecast is always the constant c. GARCH is used only for conditional variance: once a day's return has been observed, that shock can update the variance forecast for the following trading day. Repeating this procedure produces 65 one-step-ahead forecasts. Zero is used as the return benchmark, and the variance benchmark is the mean of squared returns over the previous 20 observed trading days.

DOI: 10.54254/2754-1169/2026.AB38131

5.2. Return forecasts

Compared with the zero-return benchmark, the ARMA(0,0) mean forecast lowers RMSE by 0.0029 percentage points and MAE by 0.0050 percentage points. Both gains are small. The interpretation of directional accuracy is even more important. Because c=0.0442% is positive and there are no AR or MA terms, the model predicts a positive return on every one of the 65 test days. Exactly 37 of those days have positive realized returns, so 37/65=56.9%. The 56.9% figure is therefore mostly the proportion of positive days in this test window, not evidence that the model changes its sign forecast in response to new information. The small RMSE and MAE improvements have the same source: a small positive constant is marginally closer to the realized returns than a forecast of zero. They are unrelated to the GARCH variance equation (see Table 6).

Table 6. Return forecasts compared with the zero-return benchmark

Model	RMSE	MAE	Directional accuracy
ARMA(0,0) mean forecast	0.5760	0.4190	56.9%
Zero-return benchmark	0.5789	0.4240	—

5.3. Variance forecasts

Table 7 concerns a different object from Table 6. It compares forecasts of conditional variance, not forecasts of the return itself. Student-t GARCH(1,1) has the lower MAE, while the 20-day rolling historical variance records lower MSE and QLIKE. There is therefore no across-the-board advantage for GARCH. This mixed result is plausible because MSE penalizes a few large errors more heavily, whereas MAE puts relatively more emphasis on the usual size of an error. Previous work has also shown that volatility-model rankings can change with the loss function [5]. The comparison is further complicated by the use of squared returns as a noisy proxy for latent variance [11]. In this application, the clearest role of GARCH is to provide a structured measure of conditional risk that updates over time, not to improve the sign forecast from ARMA(0,0).

Table 7. Conditional variance forecasts compared with 20-day historical variance

Model	MSE	MAE	QLIKE
Student-t GARCH(1,1)	0.43574	0.44480	-0.02800
20-day historical rolling variance	0.42923	0.44989	-0.04361

5.4. Discussion

Taken together, the mean and variance results tell a coherent story. The direction of a daily index return may be hard to predict even when volatility itself is clearly time-varying. In the estimation sample, lagged returns do not supply enough stable linear information to justify a more complicated mean equation, leaving ARMA(0,0) as the preferred specification. Its out-of-sample return forecast is consequently a fixed positive number. The variance equation behaves differently: significant α and β coefficients indicate that the size of recent shocks contains information about near-term risk. Such estimates can be useful for risk monitoring, position limits or stress analysis. They do not imply that higher volatility means the index will fall, and the 56.9% directional result should not be read as evidence of a dynamic trading signal.

DOI: 10.54254/2754-1169/2026.AB38131

6. Conclusion

The analysis uses FTSE 100 daily data from 31 May 2023 to 28 August 2026, with roughly three years assigned to estimation and the final three months reserved for testing. Returns in the estimation sample are stationary, heavy-tailed and show clear ARCH effects. BIC selects ARMA(0,0) for the conditional mean, suggesting little stable linear dependence in daily returns. For conditional variance, the retained Student-t GARCH(1,1) has statistically significant parameters. Since $\alpha+\beta=0.6379$, volatility shocks persist for a time but gradually decay.

The out-of-sample evidence is modest. The ARMA(0,0) mean forecast produces slightly lower RMSE and MAE than the zero-return benchmark, but its 56.9% directional accuracy mainly reflects the fact that the fitted mean is positive on every test day and 37 of the 65 realized returns are also positive. It is not evidence that the model dynamically forecasts market direction. For variance, Student-t GARCH(1,1) performs better only by MAE, while the 20-day rolling variance has lower MSE and QLIKE. The two equations should therefore be interpreted according to their separate roles: the mean equation offers little short-run directional information, whereas the GARCH component provides a way to describe and update conditional risk.

The study has three main limitations. It considers only one index and one relatively short test window, so the results may be sensitive to the market conditions of that period. In addition, squared returns are an imperfect and noisy measure of the latent variance that the model is trying to forecast. Finally, the main specification is symmetric and does not test structural breaks, asymmetric responses to positive and negative shocks, or additional macroeconomic predictors in a systematic way. A natural extension would be to use a longer rolling evaluation, compare forecasts against realized volatility, and examine alternatives such as GJR-GARCH, EGARCH or models with exogenous variables while keeping model selection separate from genuine out-of-sample evaluation.

References

[1] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987-1007.
[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307-327.
[3] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*(2), 347-370.
[4] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance, 48*(5), 1779-1801.
[5] Brailsford, T. J., & Faff, R. W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking & Finance, 20*(3), 419-438.
[6] Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885-905.
[7] McMillan, D. G., Speight, A. E. H., & ap Gwilym, O. (2000). Forecasting UK stock market volatility. *Applied Financial Economics, 10*(4), 435-448.
[8] Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature, 41*(2), 478-539.
[9] Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: The role of asymmetries. *International Journal of Forecasting, 21*(1), 167-183.
[10] Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1, 1)? *Journal of Applied Econometrics, 20*(7), 873-889.
[11] Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160*(1), 246-256.
[12] Liu, G. Q. (2000). Application of nonlinear GARCH models to volatility forecasting in the Chinese stock market. *Statistical Research*, (1), 49–52.

DOI: 10.54254/2754-1169/2026.AB38131

[13] Ye, Q. (2000). VaR models based on GARCH and semiparametric methods and their application to risk analysis of the Chinese stock market. *Statistical Research*, (12), 25–29.
[14] Huang, D., & Wang, H. S. (2010). Selection of estimation methods for GARCH models and their application to the Shanghai Stock Exchange Composite Index. *Journal of Applied Statistics and Management, 29*(3), 544–549.
[15] Yang, Q., & Cao, X. B. (2016). Analysis and forecasting of stock prices based on the ARMA-GARCH model. *Mathematics in Practice and Theory, 46*(6), 80–86.
[16] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.
[17] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.
[18] Investing.com. (2026). FTSE 100 historical data. *Investing.com*. https://www.investing.com/indices/uk-100-historical-data (accessed September 1, 2026).
[19] MarketWatch. (2026). FTSE 100 index (UKX) historical quotes. Historical and current end-of-day data provided by FactSet. *MarketWatch*. https://www.marketwatch.com/investing/index/ukx/download-data?countrycode=uk (accessed September 1, 2026).